AD-A250 616

REPRESENTING KNOWLEDGE AND EVIDENCE FOR DECISION

Henry E. Kyburg, Jr.
Department of Philosophy
University of Rochester
Rochester, New York 14627

Abstract

Our decisions reflect uncertainty in various ways. We take account of the uncertainty embodied in the roll of the die; we less often take account of the uncertainty of our belief that the die is fair. We need to take account of both uncertain knowledge and our knowledge of uncertainty. "Evidence" itself has been regarded as uncertain. We argue that point-valued probabilities are a poor representation of uncertainty; that we _need_ not be concerned with uncertain evidence; that interval-valued probabilities that result from knowledge of convex sets of distribution functions in reference classes (properly) include Shafer's mass functions as a special case; that these probabilities yield a plausible non-monotonic form of inference (uncertain inference, inductive inference, statistical inference); and finally that this framework provides a very nearly classical decision theory -- so far as it goes. It is unclear how global the principles (such as minimax) that go beyond the principle of maximizing expected utility are.

Science Track

keywords: evidence
             uncertainty
             decision
             non-monotonicity
             knowledge representation
             expert systems

92-13697

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE 1986	3. REPORT TYPE AND DATES COVERED Unknown

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Representing Knowledge and Evidence for Decision	DAAB10-86-C-0567

6. AUTHOR(S)

Henry E. Kyburg

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Rochester Department of Philosophy Rochester, NY 14627	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
U.S. Army CECOM Signals Warfare Directorate Vint Hill Farms Station Warrenton, VA 22186-5100	92-TRF-0026

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Statement A; Approved for public release; distribution unlimited.	

13. ABSTRACT *(Maximum 200 words)* Our decisions reflect uncertainty in various ways. We take account of the uncertainty embodied in the roll of the die; we less often take account of the uncertainty of our belief that the die is fair. We need to take account of both uncertain knowledge and our knowledge of uncertainty. "Evidence" itself has been regarded as uncertain. We argue that pointvalued probabilities are a poor representation of uncertainty; that we need not be concerned with uncertain evidence; that interval-valued probabilities that result from knowledge of convex sets of distribution fuentions in reference classes (properly) include Shafer's mass functions as a special case; that these probabilities yield a plausible non-monotonic form of inference (uncertain inference, inductive inference, statistical inference); and finally that this framework provides a very nearly classical decision theory -- so far as it goes. It is unclear how global the principles (such as minimax) that go beyond the principle of maximizing expected utility are.

14. SUBJECT TERMS	15. NUMBER OF PAGES 16
Artficial Intelligence, Data Fusion, evidence, uncertainty, decision non-monotonicity, knowledge representation, expert systems	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to **stay within the lines to meet optical scanning requirements.**

Block 1. Agency Use Only (Leave blank).

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. (If known)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD	- See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE	- See authorities.
NASA	- See Handbook NHB 2200.2.
NTIS	- Leave blank.

Block 12b. Distribution Code.

DOD	- DOD - Leave blank.
DOE	- DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA	- NASA - Leave blank.
NTIS	- NTIS - Leave blank.

Block 13. Abstract. Include a brief (Maximum 200 words) factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code (NTIS only).

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

REPRESENTING KNOWLEDGE AND EVIDENCE FOR DECISION*

One purpose -- quite a few thinkers would say the main purpose -- of seeking knowledge about the world is to enhance our ability to make sound decisions. An item of knowledge that can make no conceivable difference with regard to anything we might do would strike many as frivolous. Whether or not we want to be philosophical pragmatists in this strong sense with regard to everything we might want to enquire about, it seems a perfectly appropriate attitude to adopt toward artificial knowledge systems.

If it is granted that we are ultimately concerned with decisions, then some constraints are imposed on our measures of uncertainty at the level of decision making. If our measure of uncertainty is real valued, then it isn't hard to show that it must satisfy the classical probability axioms. For example, if an act has a real-valued utility $U(E)$ if event E obtains, and the same real-valued utility if the denial of E obtains ($U(E) = U(-E)$) then the expected utility of that act must be $U(E)$, and that must be the same as $p*U(E) + q*U(-E)$, where p and q represent the uncertainty of E and $-E$ respectively. But then we must have $p + q = 1$.[1]

There are reasons for rejecting real-valued -- i.e., strictly probabilistic -- measures of uncertainty, though not all the reasons that have been adduced for doing so are cogent. One is that these probabilities seem to embody more knowledge than they should: for example, if your beliefs are probabilistic, and you assign a probability of .1 to a drawn ball's being purple (on no evidence), and a probability of .2 to a second ball's being purple on the evidence that the first one is, and regard pairs of balls as "exchangeable"[2], then you should be 99% sure a priori that in the infinitely long run, no more than 11% of the balls will be purple. You know

beyond a shadow of a doubt (with probability .99996) on no evidence at all that no more than half will be purple. (Kyburg, 1968)

Peter Cheeseman (1985) has given a defense of classical probability, and perhaps would not find even such results as the foregoing distasteful. But it is hard to see how to defend the real-valued point of view from charges of subjectivity. Cheeseman refers to an "ideal" observer, but offers us no guidance in how to approach ideality, nor any characterization of how the ideal observer differs from the rest of us. It is therefore quite unclear what the ideal observer offers us, other than moral support: each of us is no doubt convinced that the ideal observer assigns probabilities just like himself. One man's subjectivity is another man's rational insight.[3] And there is clearly no guidance here for the construction of programs that represent probabilities.

There are other ways of representing uncertainty than by real numbers between 0 and 1. If these uncertainties are to be used in making decisions, however, they must be compatible with classical point-valued probabilities. My preference is for intervals, because they can be based on objective knowledge of distributions, and because this compatibility is demonstrable. (Kyburg, 1983)

In what follows, I will sketch the properties of interval-valued epistemic probability, and exhibit a structure for knowledge representation that allows for both uncertain inference from evidence and uncertain knowledge as a basis for decision. We need both uncertain knowledge and knowledge of uncertainty. Along the way I make some comparisons to other approaches.

1. Probability.

Probability is [4] a function from **statements** and **sets** **of** **statements** to closed subintervals of $[0,1]$. The sets of statements represent hypothetical bodies of knowledge. The idea behind $\underline{Prob}(\underline{S},\underline{K}) = [\underline{p},\underline{q}]$ is that someone whose body of knowledge is $\underline{K}$ should, ought to, have a 'degree' of belief in $\underline{S}$ characterized by the interval $[\underline{p},\underline{q}]$. The cash value of having such a 'degree' of belief is that he should not sell a ticket that returns to the purchaser \$1.00 for less than $100\underline{p}$ cents, and he should not buy such a ticket for more than $100\underline{q}$ cents. The relation in question is construed as a purely objective, logical relation.

Every probability can be based on **knowledge** of statistical distributions or relative frequencies, since statements known to have the same truth value receive the same probability, and every such equivalence class of statements (we can show) contains some statements of the appropriate form. This statistical knowledge may be both uncertain and approximate (we may be practically sure betweteen 30% and 40% of the balls are black), but it is objective in the sense that any two people having the same evidence _should_ have the same knowledge.

Classical point-valued probabilities constitute a special case, corresponding to the extreme hypothetical (and unrealistic) case in which $\underline{X}$ embodies exact statistical knowledge.

The connection between statements and frequencies is given by a set of formal procedures for finding the right reference class for a given statement. The reference set may be multi-dimensional -- the set of urns, each paired with the set of draws made from it. It may be only "accidentally" related to sentence -- as when we predict the act of someone who makes a choice on the basis of a coin toss. What is the right reference

class for a given statement S depends (formally and objectively) on what is in K, our body of knowledge. In some cases we can implement a procedure for finding the right reference class. (Loui, forthcoming.2)

It is natural to suppose that statistical knowledge in K is represented by the attribution to each reference set of a convex set of distributions -- for example we have every reason in the world to suppose that heads among coin-tosses in general is nearly binomial, with a parameter close to a half. (We have no reason to suppose that the parameter has the real value 0.49999...). Or we may have good reason to believe that two quantities are uncorrelated in their joint distribution. Or that we can rule out certain classes of extreme distributions. We can know of a certain bent coin that heads will be binomially distributed in sequences of its tosses, with a parameter p at least equal to a half.

Henceforth, we assume convexity. Here are some immediate results:[5]

(1) if $\underline{Prob}$ $(\underline{S},\underline{K})$ = $[\underline{p},\underline{q}]$ then $\underline{Prob}(-\underline{S},\underline{K})$ = $[1-\underline{q},1-\underline{p}]$.

(2) if $-$ $(\underline{S}$ & $\underline{T})$ is in $\underline{K}$, and $\underline{P}(\underline{S})$ = $[\underline{p}1,\underline{q}1]$ and $\underline{P}(\underline{T})$ = $[\underline{p}2,\underline{q}2]$ and and $\underline{P}(\underline{T}$ $\vee$ $\underline{S})$ = $[\underline{p},\underline{q}]$, then there are numbers in $[\underline{p}1,\underline{q}1]$ and $[\underline{p}2,\underline{q}2]$ whose sum is in $[\underline{p},\underline{q}]$. To see that $[\underline{p},\underline{q}]$ can be a proper subset of $[\underline{p}1 + \underline{p}2,\underline{q}1 + \underline{q}2]$, consider a die that you know to be biassed toward the one at the expense of the two, or toward the two at the expense of the one. Reasonable probability for the disjunction, "one or two" would be very close to 1/3, even though the reasonable probabilities for the one and the two would be significantly spread above and below 1/6.

(3) We can show that: given any finite set of sentences, $\underline{Si}$, and a body of knowledge $\underline{K}$, there exists a Bayesain function $\underline{B}$, satisfying the classical probability axioms, such that for every sentence $\underline{S}$ in $\underline{Si}$, $\underline{B}(\underline{S})$ $\in$ $\underline{Prob}(\underline{S},\underline{K})$.

(4) Let $\underline{KE}$ be the body of knowledge obtained from $\underline{K}$ when evidence $\underline{E}$ is

added to $\underline{K}$. If $\underline{E}$ is among the finite set of sentences in question, then there may be $\underline{no}$ Bayesian function $\underline{B}$ satisfying both $\underline{B}(\underline{S}) \in \underline{Prob}(\underline{S},\underline{K})$ and $\underline{B}(\underline{S}/\underline{E}) \in \underline{Prob}(\underline{S},\underline{KE})$: classical conditionalization is not the only way of updating probabilities.[6]

(5) There are non-trivial cases in which algorithms for computing probabilities -- i.e., for picking the right reference class -- have been provided. (Loui, forthcoming.2)

2. Updating.

A problem that has attracted a lot of attention is the problem of updating probabilities in the light of new evidence. A related problem is that of dealing with "uncertain" evidence.[7] The problem of uncertain evidence can be avoided by mechanical procedures in two well known formalisms. From a strictly Bayesian point of view, updating should take place by Jeffrey's rule: $\underline{P}'(\underline{H}) = \underline{P}(\underline{H}/\underline{E}) * \underline{P}'(\underline{E}) + \underline{P}(\underline{H}/-\underline{E}) * \underline{P}'(-\underline{E})$ (Jeffrey, 1965). The rule is not uncontroversial (Levi, 1967), but in those cases where it seems plausible, we can achieve the same result by conditioning on a piece of "certain" evidence that we expand our algebra to accommodate. Similarly, it has been shown that the same trick will work with Glenn Shafer's well known mathematical theory of evidence (Shafer, 1976): we can mechanically replace general combination of support functions, so long as the evidence can be represented by a seperable support function, by Dempster conditioning -- Shafer's analog to Bayesian conditionalization. (Kyburg, forthcoming.1)

The relation between Shafer's theory and the system of probability just outlined is interesting. Let θ be a possibility space, with support function $\underline{s}$ defined on it. Shafer also defines a plausibility function $\underline{t}$: for every subset $\underline{S}$ of θ, $\underline{t}(\underline{S}) = 1 - \underline{s}(\theta - \underline{S})$. Of course subsets of a

possibility space correspond exactly to propositions, and we can construct a convex set of probability functions over these propositions such that the minimum and maximum probabilities assigned to a proposition are exactly the support and plausibility of the corresponding subset of 0. (Kyburg, forthcoming.1)

But the converse doesn't hold. Consider a compound experiment consisting of either (1) tossing a fair coin twice, or (2) drawing a coin from a bag containing 40% two-headed and 60% two-tailed coins and tossing it twice. The two alternatives are performed in some unknown ratio p. Let A be the event that the first toss lands heads, and B the event that the second toss lands tails. The representation by a convex set of probability functions is straight-forward:

$$P(\underline{TT}) = \underline{p}/4 + 0.6(1-\underline{p})$$

$$P(\underline{TH}) = \underline{p}/4$$

$$P(\underline{HT}) = \underline{p}/4$$

$$P(\underline{TT}) = \underline{p}/4 + 0.4(1-\underline{p})$$

The convex set of probability measures over the sample space is just the set of these values for $\underline{p} \in [0,1]$. Let this set be $\underline{SP}$. $\underline{P}^*(\underline{S}) = \min \{ \underline{P}(\underline{S}) : \underline{P} \in \underline{SP} \}$ is not a support function, by theorem 2.1 of (Shafer, 1976). (Kyburg, forthcoming.1)

Finally, let $\underline{CP}(\underline{e})$ be the set of probability functions resulting from conditionalizing the members of $\underline{P}$ on $\underline{e}$. That is, if $\underline{p}$ belongs to $\underline{P}$, then the function $\underline{p}(\underline{x}/\underline{e}) = \underline{p}(\underline{x}\&\underline{e})/\underline{p}(\underline{e})$ defined for every sentence $\underline{x}$ in the original algebra will belong to $\underline{CP}(\underline{e})$.[8] $\underline{CP}(\underline{e})$ is a convex set of classical probability functions. Let $\underline{CPle}$ be the corresponding lower-probability function, and $\underline{CPue}$ the corresponding upper-probability functin. (Neither are probability functions -- hence the hyphens are not accidental.) Let

DP$_s$e be the support function obtained from the support function $\underline{s}$
corresponding to $\underline{P}$ by Dempster conditioning -- i.e., Dempster's rule of
combination applied to the case where $\underline{e}$ receives unit support. Let
DP$\underline{p}$e be the corresponding plausiblity function. Then

$$CPle(\underline{s}) \leq DPse(\underline{s}) \leq DPpe(\underline{s}) \leq CPue(\underline{s})$$

Inequality holds unless certain measures on subsets have the value 0. When
it comes to updating probabilities relative to evidence, Shafer's procedure
exaggerates the impact of evidence beyond its Bayesian import. (Kyburg,
forthcoming.1)

But we can also specify exactly the conditions under which this form of
updating agrees with convex Bayesian conditionalization. If these
conditions are satisfied, then it makes sense to follw the Dempster-Shafer
formalism, especially when it is computationally simpler.

Bayesian conditionalization is not always the right way of updating
probabilities, however. A situation in which Bayesian conditionalization
whould be given up appears in (Kyburg, forthcoming.2)

3. Uncertain Knowledge

One problem that Bayesian and other approaches to uncertainty have is
that there is no formal way of representing the acquisition of knowledge.
We can represent the _having_ of knowledge (by the assignment of probability 1
to the item), but since there is no way in which $\underline{P}(\underline{S}/\underline{E})$ can be 1 unless $\underline{P}(\underline{S})$
is already one, conditionalization doesn't get us knowledge. This has been
noticed, of course; Cheeseman (1985, p. 1008) simply says, "A reasonable
compromise is to treat propositions whose probability is close to 0 or 1 as
if they are known with certainty...." But of course it is well known that
this cannot be done generally: the conjunction of a number of certainties is

a certainty, but the conjunction of a large enough number of "reasonable certainties" in Cheeseman's sense is what he would have to consider an impossibility.[9]

McCarthy and Hayes (1969) are seduced into following this primrose path, when they suggest (p. 489) "If $\theta_1, \theta_2, \ldots, \theta_n \vdash \theta$ is a possible deduction, then <u>probably</u>$(\theta_1), \ldots, $<u>probably</u>$(\theta_n) \vdash $<u>probably</u>$(\theta)$ is also a possible deduction." This is clearly ruled out, on our scheme -- and even <u>acceptable</u>$(\theta_1), \ldots, $<u>acceptable</u>$(\theta_2) \vdash $<u>acceptable</u>$(\theta)$ is ruled out as a consequent of the logical conditional. Many philosophers, of course, have taken this for granted -- but if we are to formalize uncertain inference at all, we must somehow accommodate sets of conflicting statements. <u>Purely</u> probabilistic rules of inference do this easily.

We can accommodate Cheeseman's intuition that we should accept what is practically certain by considering <u>two</u> sets of sentences in the representation of knowledge. One of them we will call the evidential corpus, and denote by <u>Ke</u>; the other we will call the corpus of practical certainties, and denote by <u>Kp</u>.

We will accept sentences into <u>Kp</u> if and only if their probability relative to <u>Ke</u> is greater than <u>p</u>. The conjunction of two statements that appear in <u>Kp</u> will also appear in <u>Kp</u> <u>only</u> if the conjunction itself is probable enough relative to <u>Ke</u>. Thus <u>Kp</u> will <u>not</u> be deductively closed, though we can prove that if a statement <u>S</u> appears in <u>Kp</u>, and <u>S</u> entails <u>T</u>, <u>T</u> will also appear there. This reflects a natural feature of human inference: we must have reason, not only to accept each premise in a complex argument, but to accept the <u>conjunction</u> of the premises, in order to be confident of the conclusion.

We have a picture that looks like this:

```
**********
*  Ke   *
*        *
**********
```

$$\Downarrow$$

Uncertain inference: $\underline{S} \in \underline{Kp}$ iff
$\underline{Prob(\underline{S},Ke)} \geq \underline{p}$.

```
**********
*        *
*  Kp   *
**********
```

It is relative to $\underline{Kp}$, the practical corpus, that we make our
(practical) decisions. It is thus the (convex sets of) distributions --
including conditional distributions -- embodied in that set of statements
that we use in our decision theory.

But there are questions. What is the value of $\underline{p}$ that we are taking as
practical certainty? How do statements get in $\underline{Ke}$? What is the decision
theory that goes with this kind of structure?

Let us first consider the value of $\underline{p}$. Suppose the widest range of
stakes we can come up with is 99:1. For example, Sam and Sally are going to
bet on some event, each has $100, and neither has any change. Then a
probability value falling outside the range of $[.01,.99]$ would be useless as
a betting guide. A probability less than .01 would (in this context) amount
to a practical impossibility; one greater than .99 would amount to a
practical certainty.

The range of stakes can determine the level of "practical certainty" $\underline{p}$.
What counts as practical certainty depends on context, but in an explicit
way: it depends on what's at stake. This idea is developed in (Kyburg,
forthcoming.2).

How do statements qualify as evidence in $\underline{Ke}$? Not by being "certain."

It can be argued that anything that was really incorrigible would have to be devoid of empirical content.[10] (The worry about uncertain evidence is not misplaced; it's just misconstrued.) One typical form of evidence statement is this: "The length of x is d + r meters." Whatever our readings, these statements are not "certain" -- they admit of error. The same is true of all ordinary observation statements.

So a statement gets into Ke by having a low probability of being in error; equally, by having a high probability (at least e) of being veridical. How high? In virtue of the fact that conjunctions of pairs of statements in Ke appear in Kp, it seems plausible to take $e = (p)^{1/2}$. For a number of technical reasons (Kyburg, 1984) it turns out to be best to construe the corpus containing the theory of error as metalinguistic. This is as one might think: after all, the theory of error concerns the relation between readings -- e.g. numerals written in laboratory books -- and values: the real quantities characterizing things in the real world. For present purposes we need note only that this is not the begining of an infinite regress. We can maintain objectivity; we can avoid "presuppositions" and other unjustified assumptions.

4. Decision.

It has been objected (Seidenfeld, 1979) that there is no decision theory that is tailored to Shafer's theory of evidential support. Indeed, it is pretty clear that support functions alone would conflict with expected utility. On the other hand, the reduction to convex sets of distributions does show that we can have very nearly a normal decision theory using Shafer's system. In computing the value of an act, we need to consider not only the support assigned to various states of affairs (corresponding to lower probabilities), but also the plausibilities -- corresponding to upper

probabilities.)

This is true for the more general convex set representation: We can construct an _interval_ of expected utility for each act. A natural reinterpretation of the principle of dominance would take an alternative $a1$ to dominate an alternative $a2$ whenever, for every possible frequency distribution, the expectation of $a1$ is greater than the expectation of $a2$.

This eliminates some alternatives, but in general there will be a number of courses of action that are not eliminated. What we do here is another matter, one which is certainly worthy of further study.[11] But it seems natural that minimax and minimax regret strategies are appropriate candidates for consideration under some conditions. There may well be others, such a satisficing. And it may even be that the guidance provided by the motto: eliminate dominated alternatives, is as far as rationality alone takes us. Further pruning may depend on constraints that are local to the individual decision problems.[12]

5. The Structure of Knowledge.

Were we to deal explicitly with our theory of error and its source, we would have a complex structure consisting of four sets of sentences in two distinct languages.[13] But for ordinary decision theoretic purposes there are just two sets of statements with which we need to be concerned Kp and Ke. Evidence enters Ke when it is dependable enough, and Ke in turn determines the practical certainties of Kp. This renders the process of uncertain inference by which _any_ statement gets into Kp automatically non-monotonic. As the contents of the evidential corpus Ke changes, Kp may change, contract, or expand. What is practically certain at one point may cease to be practically certain in the light of new evidence, and in fact in the light of new evidence may become _evidently_ false.[14]

Another feature of the relation between the evidential corpus and the practical corpus is that sentences in the evidential corpus are inherited by the practical corpus. The practical corpus is thus an expansion of the evidential corpus; but it is crucial to keep the two corpora distinct. If a sentence were to be added to the evidential corpus when it got a high probability relative to the evidential corpus, it could never be eliminated: it would henceforth always have probability one relative to that evidential corpus. The separation of the practical and evidential corpus is <u>required</u> to preserve the non-monotonicity of uncertain inference.

The decision maker need be concerned directly only with the contents of <u>Kp</u> -- that is what determines the (objective, frequency-based) probability of the alternatives he must choose between. But he <u>may</u> be led to worry about the contents of <u>Kp</u>. What is there depends on the weight of the combined evidence concerning it. This evidence is embodied in <u>Ke</u> and the mode of combination flows from the definition of probability.

The scheme outlined does not give us a complete decision theory such as we would get from a subjective Bayesian approach, but it may take us as far as rationality can take us. The role of epistemological probability in decision theory is supported by the theorem that for any finite set of sentences there is a Bayesian belief function that <u>fits</u> the epistemological probability intervals. Thus uncertain knowledge and knowledge of uncertainty both find their place.

1. It is this line of attack that lies behind the subjectivist approach to probability established independently by F. P. Ramsey (1930) and Bruno de Finetti (1937) and rendered respectable by L. J. Savage (1954).

2. If "Pi" is "Draw number i yields a purple ball," this is just to say that for $i \rightleftharpoons j$ Prob(Pi) and Prob(Pi & Pj) do not depend on the values of i and j.

3. There is a tradition, represented by H. Jeffreys (1939), R. Carnap (1950), and most recently E. T. Jaynes (1982), according to which the subjectivity of precise probability assignments can be eliminated by finding general principles for assigning probabilities to the statements of a given language. But as Seidenfeld (forthcoming) has shown, there are serious difficulties with the Maximum Entropy program even beyond the fact that this approach just pushes the arbitrariness into the choice of a language or classification.

4. Of course this is just one opinion among many as to what probability "is". But I would hardly hold it if I did not think it correct.

5. roofs may be found in (Kyburg, 1961), (Kyburg, 1974) and (Kyburg, 1983).

6. Counterillustration may be found in (Kyburg, forthcoming.3).

7. Simply as examples: (Duda, Hart, and Nilsson, 1976), (Garvey, Lowrance, and Fishler 1981), (Pearl, 1985), (Lowrence, 1982), Quinlan, 1982).

8. We assume $p(e) > 0$ for every $p \in P$; we also assume that there is a support function s matching P.

9. This is the lottery paradox, first appearing in (Kyburg, 1961).

10. One normally believes one's own eyes, but one knows that hallucinations

<u>do</u> occur. It is hard to imagine any observational statements whose veridicality could not be impugned by some imaginable course of subsequent observations. Perhaps this is not true of phenomenological reports: "Red patch here now." But I suspect these have no useful content.

11. See Levi (1980) for a highly developed form of this approach.

12. Or perhaps this whole approach is wrong-headed. For the development of an alternative, see (Loui, forthcoming.1).

13. Viz.: the practical corpus <u>Kp</u>, the evidential corpus <u>Ke</u>, the evidential metacorpus <u>MKe</u>, and the <u>a priori</u> metacorpus <u>MKa</u> containing observational records and linguistic conventions.

14. Note that in a strict sense, <u>Kp</u> need not even be consistent -- that is, its deductive closure may be inconsistent in the ordinary sense. This is illustrated by the lottery alluded to.

Carnap, Rudolf (1950): The Logical Foundations of Probability, University of
    Chicago Press, Chicago.

Cheeseman, Peter (1985): "In Defense of Probability," IJCAI L985, II, pp.
    1002-1009.

Duda, Hart, and Nilsson, (1976): "Subjective Bayesian Methods for Rule-
    Based Inference Systems," Proceedings of the National Computer
    Conference 45, pp. 1075-1082.

Finetti, Bruno (1937): "La Prevision: Ses Lois Logiques, Ses Sources
Subjectives," Annales de L'Institute Henry Poincare 7, 1937, pp. 1-68.

Garvey, Lowrance, and Fishler, (1981): "An Inference Technique for
    Integrating Knowledge from Disparate Sources," Proceedings IJCAI 7, PP.
    319-325.

Jaynes, E.T. (1982): "On the Rationale of Maximum Entropy Methods,"
    Proceedings of the IEEE 70, pp. 939-952.

Jeffrey, Richard (1965): The Logic of Decision, McGraw-Hill, New York.

Jeffreys, Harold (1939): Theory of Probability, Oxford University Press,
    Oxford.

Kyburg, Henry E., Jr. (1968): "Bets and Beliefs," American Philosophical
    Quarterly 5, pp. 54-63.

_____ (1961): Probability and the Logic of Rational Belief, Wesleyan
    University Press, Middletown.

_____ (1974): The Logical Foundations of Statistical Inference, Reidel,
    Dordrecht.

_____ (1983): "The Reference Class," Philosophy of Science 50, pp. 374-
    397.

_____ (1984): Theory and Measurement, Cambridge University Press,
    Cambridge.

_____ (Forthcoming.1): "Bayesian and Non-Bayesian Evidential Updating,"
    Artificial Intelligence.

_____ (Forthcoming.2): "Full Belief."

_____ (Forthcoming.3): "The Basic Bayesian Blunder."

Levi, Isaac (1968): "Probability Kinematics," British Journal for the
    Philosophy of Science 18, pp. 197-209.

_____ (1980): The Enterprise of Knowledge, MIT Press, Cambridge.

Loui, Ronald P. (Forthcoming.1): "Interval Based Decisions for Reasoning
    Systems," Proceedings of the UCLA Workshop on Uncertainty and

_____ Probability in Artificial Intelligence, John Lemmon (ed.).

_____ (Forthcoming.2): "Computing Reference Classes."

Lowrance, John (1982): "Dependency Graph Models of Evidential Support,"
    University of Massachusetts, Amherst.

McCarthy, John, and Hayes, Patrick (1969): "Some Philosophical Problems
    from the Standpoint of Artificial Intelligence," Machine Intelligence
    4, pp. 463-502.

Pearl, Judea (1985): "Fusion, Propagation, and Structuring in Bayesian
    Networks," TR CSD-850022, UCLA, Los Angeles.

Quinlan,   (1982): "Inferno: A Cautious Approach to Uncertain Inference, A
    Rand Note," California.

Ramsey, F.P. (1931): The Foundations of Mathematics and Other Essays,
    Humanities Press, New York.

Savage, L.J. (1954): The Foundations of Statistics, John Wiley, New York.

Seidenfeld, Teddy (1979): "Statistical Evidence and Belief Functions" PSA
    1978, Asquith and Hacking (eds.).

_____ (Forthcoming): "Entropy and Uncertainty."

Shafer, Glenn (1976): A Mathematical Theory of Evidence, Princeton
    University Press, Princeton.